DOCUMENT RESUME

ED 330 385 JC 910 168

AUTHOR Boughan, Karl
TITLE A Cluster Analysis of the 1984 Entrant Survey: Implementing Geo-Demographic Marketing at P.G.C.C., Part III. Market Analysis MA91-6.
INSTITUTION Prince George's Community Coll., Largo, MD. Office of Institutional Research and Analysis.
PUB DATE Mar 91
NOTE 14p.; For related documents, see ED 325 182 and JC 910 166-167.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Associate Degrees; *Cluster Analysis; College Transfer Students; Community Colleges; Credit Courses; *Educational Status Comparison; Followup Studies; Institutional Research; Outcomes of Education; Research Problems; *Socioeconomic Influences; *Socioeconomic Status; Student Educational Objectives; Two Year Colleges; *Two Year College Students
IDENTIFIERS Prince Georges Community College MD

ABSTRACT

Prince George's Community College (PGCC), Maryland, utilizes a geo-demographic marketing model in which neighborhoods within the college's service area are grouped into natural socioeconomic "clusters." In an effort to track the outcomes of PGCC students 4 years after entrance and to compare outcomes by the different socioeconomic clusters of students, a study was conducted of 414 students who were first-time credit students in the fall of 1984. Data were obtained from the Maryland State Board of Community Colleges which had administered two surveys to this student cohort, once at entry to PGCC in 1984 and again in 1988 after 4 years. Students were identified as belonging to one of 13 socioeconomic clusters and were analyzed with regard to 6 academic outcome indicators: "Achievers" (degree attainers); "Transfer-Seekers"; "Degree-Seekers"; "Muddlers-Through" (transfer to four-year institution without degree award); "Strugglers" (still enrolled at PGCC); and "Upgraders" (no-award, no-transfer, no-longer-attending). Selected findings included the following: (1) "Achievers" were from predominantly white, upscale groups; (2) "Transfer-Seekers" included predominantly black or mixed-race, middle class populations, similar in outcomes to "Achievers" but with a higher rate of termination without achievement; (3) "Degree-Seekers" included mostly white middle-class groups seeking associate degrees but with little interest in baccalaureate work; (4) the "Muddlers-Through" group included a very significant minority component, many of whom were lower white-to-blue-collar workers; (5) "Strugglers" included mainly single blue-collar youths; and (6) "Upgraders" included predominately college-educated young people apparently seeking skills upgrading. Data tables are included. (GFW)

ED330385

IMPLEMENTING GEO-DEMOGRAPHIC MARKETING AT P.G.C.C.

Part III

A Cluster Analysis of the 1984 Entrant Survey

PG TRAK©

"PERMISSION TO REPRODUCE THIS MATERIAL
HAS BEEN GRANTED BY
K. M. Boughan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Karl Boughan
Prince George's Community College
Office of Institutional Research and Analysis

Market Analysis MA91-6
March 1991

JC 910168

BEST COPY AVAILABLE

PRINCE GEORGE'S COMMUNITY COLLEGE
Office of Institutional Research and Analysis

IMPLEMENTING GEO-DEMOGRAPHIC MARKETING AT P.G.C.C. Part III -- ADDITIONAL CREDIT OUTCOMES FROM THE S.B.C.C. 1984 ENTRANT SURVEY

Market Analysis MA91-6
March 1991

Introduction

In Implementing Geo-Demographic Marketing at P.G.C.C. Part I (MA91-4, December 1990), we presented an academic outcomes study based on a PG-TRAK(c)-coded unduplicated member file of all P.G.C.C. credit course-taking students 1985-1990. Its main purpose was to cluster-track academic outcomes to the full extent permitted by the roster of indicators maintained on the College's all-credit student historical file (TAB).

Unfortunately, TAB's range of variables does not extend to those marking student post-P.G.C.C. academic careers. For the study mentioned, this meant that vital data on student transfer rates by cluster, as well as other important outcome statistics dependent upon transfer career information (e.g., post-A.A. degree attainment rates and proportions of no-outcome students), could not be calculated.

This report presents the findings of a study designed to fill in some of the gaps left by Part I in our knowledge of student cluster academic outcomes. This new research was not grounded on the clusterized 1985-1990 TAB sample but on the P.G.C.C. segment of the Maryland State Board of Community College's massive 1984 Entrant Survey sample.

The S.B.C.C.'s objective was specifically to get around the lamentable lack of "hard" transcript-based transfer-related data, a problem for all seventeen community colleges in the state. It did so in the spring of 1988 by going directly to all Maryland 1984 first-time community college credit students with a transfer item-loaded questionnaire.

We have already utilized this 1984-1988 all-college data set in a study of state-wide and P.G.C.C.-wide academic outcomes. (See Award Attainers at Maryland Community Colleges, RB91-1, July 1990.) In this case, we sifted out the P.G.C.C. sub-sample and cluster-flagged our respondents by running a survey sub-sample-to-clusterized TAB file match, to produce a cohort of P.G.C.C. students whose full four-year community college-through-transfer school career could be tracked by cluster.

Special Methodological Considerations

In switching to a new data base, we encountered several technical considerations and problems.

Data Base Differences. The first one was that, although the two P.G.C.C. credit student samples were similar, they were not identical. The S.B.C.C. sample was designed to generate outcome assessment after four years; the TAB sample featured a five-year interval and was not drawn specifically for outcomes assessment purposes. Furthermore, the two time spans imperfectly overlapped -- 1984-1988 to 1985-1990, respectively. Finally, S.B.C.C. respondents were all first-time/any college students as of 1984. On the TAB subfile side, the situation is considerably muddier, for no restriction was placed limiting sample members to 1985 first-timers.

All of the above means that the outcome results of the S.B.C.C. data analysis can be read back onto our recent credit student body only in a general way; and in particular the S.B.C.C. cluster-by-cluster graduation percentages will fall short of those generated by the TAB file analysis, sometimes considerably. (This is only to be expected since the latter restricts outcome assessment to a single four-year interval, while the former yields graduation rates based on five to possibly thirty-three years of study.) This weak sample-to-sample parallelism is a disadvantage.

On the other hand, using the S.B.C.C. data has this very great virtue when study focus is on academic outcome, as here -- the 1984 Entrant Survey was specifically constructed with outcome assessment in mind. That is the reason in the first place for limiting the assessment period to a single interval and students to first-timers: these are the very conditions for a clear test. Therefore, if study design is the only consideration, the S.B.C.C. data is the more appropriate and ought to yield superior outcome statistics.

Sample Size and Stability of Estimates. A more vital methodological problem revolves around the size of the P.G.C.C. sub-sample in the total S.B.C.C. data file. The original sub-sample consisted of 624 College respondents, a number more than adequate to gauge accurate divisions of response for the whole group (assuming sample representativeness). But our objective required the analysis not only of College sub-sample responses but also of responses for twenty-four cluster sub-sub-samples. This gives us, by simple arithmetic, a mean sub-sub-sample size of only 26! And the problem becomes even worse when results of sample loss due to cluster-encoding are factored in -- College sample N=414, mean cluster sub-sample n=17.

According to tested statistical tradition, no percentages should ever be calculated on sample bases smaller than 20 observations; estimates are simply too unstable, too sensitive to

possible random fluctuations of only one or two cases. The average cluster sub-sample figure already fell below 20, and many actual cluster "n's" dropped way off the minimum standard for percentaging.

We attempted to deal with this problem by aggregating clusters into larger "super-clusters." We employed three rules-of-thumb: First, if a cluster's "n" exceeded or came near 20, we retained that cluster as a full unit of analysis. But, if it failed the test, then an attempt was made to group it with one or more other clusters based on geo-demographic similarity -- which would result in some loss of lifestyle accuracy but at least would preserve respondent data. Third, if "super-clustering" failed due to high cluster distinctiveness, that cluster and its respondents were simply dropped from further analysis. The results can be seen in Table 1 below:

TABLE 1. Super-Cluster Formation:
Sample Characteristics and Components

SUPER-CLUSTER	UNWTD N	WTD %	CLUSTER COMPONENTS
A-Country Club	36	7.4	Cnt Clb-1
B-Exurban Dream	41	8.3	Exrb Dr-2
C-Aging White Collr	43	8.0	Ag Affl-3,Bwy Hvn-5
D-Rainbow Manors	26	5.5	Rbw Mnr-6
E-Bright Beginnings	36	11.6	Br Begs-8
F-Rapid Development	47	12.0	Homstdr-9,Bm Twn-20,Cnt Bl-23
G-New Collars	21	4.7	Nw Col-10
H-Levittown, P.G.	22	5.6	Levitn-15
I-Minority Wh Collr	45	11.0	Srg Mn-11,Mn Row-16,Emg Mn-17
J-Middle America	18	3.2	Mid Am-18
K-Old Timers	18	3.1	Old-Tm-19
L-Downtown, P.G.	21	8.0	DntnPG-22
M-Minority Bl Collr	31	9.8	BC Blk-21,Cty Ln-24
--ENTIRE SAMPLE----	414	100.00	

NOTE: Clusterized SBCC 1984 Entry Survey PGCC sample failed to capture any respondents from Sophisticate Mix-4, Government Mix-7 or Dormitories Plus-13; also, Clusters 12 (Ft. George) and 14 (Bohemian Mix) registered only four respondents each, and as they were demographically uncombinable with other clusters these were dropped from further study consideration except that their data are reflected in Whole Sample statistics.

Thirteen clusters and super-clusters emerged after these steps were taken. The super-cluster exigency was needed in four instances (grouping upper middle class "empty-nest," middle class black, blue collar black, and fast-developing but still underpopulated rural clusters). Three clusters -- very small in any case in terms of student body population -- failed to contribute a single S.B.C.C./P.G.C.C. respondent: Sophisticate Mix-4 (.1 percent of student body), Government Mix-7 (.8 percent) and Dormitories Plus-13 (almost 0 percent). Two other slightly more student body-represented clusters -- Ft. George-12 (2.0 percent) and Bohemian Mix-14 (1.9 percent) -- each generated a mere four respondents, and since their relative lifestyle uniqueness precluded joining them either to other clusters or each other, they were removed from further consideration.

Sample Bias and Re-Weighting. In Part I the systematic distortion of mail survey samples, removal of sample bias by means of sample weighting, and the specific weighting techniques we employed in that study to restore S.B.C.C. sample representativeness is fully discussed. Unfortunately, simply re-using the original weighting scheme would not prove effective here given the one-third reduction of sample size brought about by cluster encoding and the consequent re-shifting of the sample's demographic and academic characteristics. Therefore, using the same weighting factors as in the earlier study a new weighting variable was specially created.

The re-weighting step, however, proved insufficient to remove damaging sample bias entirely -- often the case when radical sample size reduction occurs. This became evident when we compared outcome percentages of the weighted whole original sample with those of the re-weighted whole reduced sample: relatively small but significant differences were found. Since the original percentages were, by statistical definition, the more accurate, some step had to be taken to bring the new percentages more in line with them were our cluster outcome estimates to have any accuracy. That step was the forced re-proportioning of cluster outcome percentages by systematically multiplying them by coefficients representing the whole sample over- or under-proportion in each outcome case.

For example, the original whole sample estimate of four-year transfer rate was 26.96 percent whereas the present whole sample produced an estimate of only 21.05 percent. To restore the 26.96 figure requires a multiplication of 21.05 by a coefficient equaling 1.281 (26.96/21.05). Our method in essence called for applying this 1.281 coefficient not only to the new whole sample estimate, but also to each cluster sub-sample estimate. (For the details of our re-proportioning method, see Appendix.)

Concluding Methodological Caveats. It should be apparent by now, with all that has been said concerning sample differences and the need for a great deal of "massaging" to overcome data

distortions, that the actual findings should be viewed more as suggestive than absolute. What follows is a summary of warnings for interpreters of the actual findings:

1. Estimates from this study, strictly speaking, reflect only the four-year academic attainments of P.G.C.C. first-time/any college credit students who began in the Fall of 1984.

2. It cannot be strongly enough emphasized that the outcome percentages generated by this study are estimates only. The self-selecting nature of the respondent sample inherent in the mail survey approach inevitably results in high sample unrepresentativeness, which can never be completely rectified by means of sample weighting. Furthermore, attempts to improve weighted sample representativeness by re-proportioning based on the known values of criterion variables, while bettering estimate accuracy in general, adds uncertainty to any individual estimate.

3. Except for community college award attainment, outcome rates are based only upon the reports of respondents, not upon the "hard" data of transcript files. Inevitably, some unknown proportion of respondents will either lie about or mis-remember the details and outcomes of their academic careers.

Findings: P.G.C.C. Academic Outcomes by Cluster

Table 2 (page 6) provides a comprehensive listing of all study outcome estimates by super-cluster, ordered roughly by socio-economic status. The six key outcome indicators selected for presentation were:

- P.G.C.C. Degree Attainment - A.A., Certificate or L-O-R
- Reported Transfer to 4-Year College or University
- Overall "Academic Achievement" - Award and/or Transfer 4-Yr
- "Pass-Through" Rate - Transferring without Award
- "Pipeline" Rate - Still at P.G.C.C. (Spring '88)
- "Pass-Out" Rate - No Longer Attending/No Award/No Transfer

Both actual percentage estimates are shown for each super-cluster, as well as corresponding index scores (100 x cluster % / whole sample %). (In addition, the Appendix provides at table giving percentage and index scores for each super-cluster and indicator in super-cluster rank-order.)

TABLE 2. Selected Academic Outcome Indicators by Super-Cluster*

SUPER-CLUSTER	% AA/ CERT.	% 4-YR. TRNSF.	% AWRD/ TRNSF.	% PASS- THRU	% PIPE- LINE	% PASS- OUT
A-Country Club	11	19	22	12	2	75
B-Exurban Dream	9	38	39	42	11	38
C-Aging White Collr	10	31	35	39	14	36
D-Rainbow Manors	8	30	33	34	7	52
E-Bright Beginnings	8	17	20	12	33	47
F-Rapid Development	10	37	40	33	7	50
G-New Collars	7	34	35	31	0	63
H-Levittown, P.G.	13	9	18	5	7	75
I-Minority Wh Collr	4	36	39	36	7	53
J-Middle America	15	26	32	17	0	68
K-Old Timers	7	21	26	19	3	71
L-Downtown, P.G.	4	8	12	8	4	84
M-Minority Bl Collr	8	25	30	22	7	63
--ENTIRE SAMPLE----	9	27	31	25	9	57

SUPER-CLUSTER	Idx** AA/ CERT.	Idx 4-YR. TRNSF.	Idx AWRD/ TRNSF.	Idx PASS- THRU	Idx PIPE- LINE	Idx PASS- OUT
A-Country Club	127	70	73	45	26	133
B-Exurban Dream	110	141	128	167	112	67
C-Aging White Collr	121	116	113	155	147	64
D-Rainbow Manors	89	110	107	133	70	92
E-Bright Beginnings	92	65	65	47	351	83
F-Rapid Development	112	136	132	129	79	89
G-New Collars	77	127	113	120	0	111
H-Levittown, P.G.	155	32	58	18	76	133
I-Minority Wh Collr	51	133	127	140	72	94
J-Middle America	174	96	105	68	0	120
K-Old Timers	85	79	84	73	29	126
L-Downtown, P.G.	46	31	40	33	43	148
M-Minority Bl Collr	93	94	98	87	70	112
--ENTIRE SAMPLE----	100	100	100	100	100	100

INDICATOR KEY:

- AA/CERT = A.A., Certificate or Letter of Recognition
- 4-YR TRNSF = Claimed transfer to 4-yr institution
- AWARD/TRNSF = Either PGCC Award, 4-yr Transfer or both
- PASSTHRU = Transferred without graduating
- PIPELINE = Still at PGCC as of Spring 1988
- PASSOUT = No longer at PGCC, no grad., no transfer

* All figures have been rounded for table display

** Index = 100 x (Super-Cluster % / Entire Sample %)

It is a bit difficult to grasp the cluster outcome patterns displayed in Table 2, given the large amount of data it contains. Therefore, we have attempted to simplify this task by re-organizing the data in the table's bottom half, grouping super-clusters according to index score similarities across all six indicators:

TABLE 3. Super-Clusters Blocked by Outcome Similarity
(Mean Cluster Index Scores within Blocks)

OUTCOME BLOCKS		AA/ CERT.	4-YR. TRNSF.	AWRD/ TRNSF.	PASS-THRU	PIPE-LINE	PASS-OUT
I.	"Achievers"	115	128	120	167	129	65
II.	"Transfer-Seekers"	84	126	122	139	74	92
III.	"Degree-Seekers"	152	66	79	45	34	128
IV.	"Muddlers-Thru"	85	100	99	97	33	116
V.	"Strugglers"	46	31	40	34	43	148
VI.	"Upgraders"	92	65	65	49	351	83

BLOCK KEY:
I. B-Xrb Drm C-Ag WCol
II. D-Rbw Mnr F-Rap Dev I-Min WCl
III. A-Cnt Clb H-Levittn J-Mid Amr
IV. G-New Col K-OldTmr M-Min BC
V. L-Downtwn
VI. E-Brt Beg

Six broad outcome patterns emerge:

I. "Achievers" (B-Exurban Dream, C-Aging White Collar). Two upscale, predominantly white clusters at the end of their child-rearing family cycle form Block I. It evinces an academic outcome pattern of striving and success -- above average rates of study persistence ("Pipeline") and either graduation, transfer or both. Also significant is the strong Achiever tendency to skip over community college graduation and move on directly to four-year schools. However, P.G.C.C. degree attainment is still disproportionate here.

II. "Transfer-Seekers" (D-Rainbow Manors, F-Rapid Development, I-Minority White Collar). Block II is composed of three basically middle class super-clusters featuring predominantly black or mixed-race populations and known for exhibiting a high level of socio-economic ambition. Block II's outcome pattern is quite like that of the "Achievers" above, except that its graduation-to-passthrough ratio is considerably lower (indicating even a stronger level of impatience to move on). On the other hand, persistence (the "Pipeline" rate) is much in evidence and the "pass-out" rate, in part an index of termination without achievement, is significant higher although still below average.

III. "Degree-Seekers" (A-Country Club, H-Levittown, P.G., J-

Middle America). Block III pulls together three mostly white middle class clusters, two of which are distinctly lower white collar/upper blue collar in occupational character and modest in levels of family income and adult education exhibited -- in short, "Middle American." The "Degree-Seekers" seem, by their outcome scores, to be interested in the immediate achievement of a community college degree; the possibility of going on towards a BA/BS apparently leaves them cold. Their strategy appears to be "get in-get certified-get out" -- for despite their high rate of graduation, long-term study is almost absent among them. This pattern suggests to us a very pragmatic, probably occupational orientation.

IV. "Muddlers-Through" (G-New Collars, K-Old Timers, M-Minority Blue Collar). Made up out of three lower white-to-blue collar clusters with a very significant minority component, Block IV exhibits an uninspiring, mostly middling or a little below set of outcome indicator scores.

V. "Strugglers" (L-Downtown, P.G.). Block V represents a single cluster inhabited mainly by single, blue collar youths. It registers very poorly on all positive academic achievement indicators, shows little persistence, and a very high "pass-out" rate -- definitely the "at risk" segment of our student body.

VI. "Upgraders" (E-Bright Beginnings). Block VI houses a solitary but very important cluster -- Bright Beginnings, the single largest lifestyle component of our student body (11 percent) and sociological very distinctive. Cluster 8s are mainly already college-educated young people at the start of their professional careers. This probably explains its weird outcome pattern -- sub-average degree and transfer attainment (unneeded) but also low rate of "passing-out" and a truly extraordinary rate of continuing enrollment. To us, this suggests that professionally mobile Cluster 8s mostly use P.G.C.C. credit courses for job skills upgrading, and do so habitually over time.

Conclusions

The PG-TRAK(c) lifestyle clusters continue to show their worth in illuminating P.G.C.C.'s student behavior, in this case by lending themselves to a re-analysis of the College component of a state-supplied survey into post-community college academic activity. This has allowed us to fill in many gaps in our knowledge of P.G.C.C. patterns of academic outcome in a fashion which should aid in both academic planning and enrollment management.

A * P * P * E * N * D * I * X

I. Super-Clusters Rank-Ordered by Outcome Rates

II. Details of Re-Proportioning Methodology

TABLE 3. Super-Clusters Ranked by Academic Outcome Indicators

PGCC AWARD (AA/CERT.)	%	IDX
J-Middle America	15	174
H-Levittown, P.G.	13	155
A-Country Club	11	127
C-Aging White Collr	10	121
F-Rapid Development	10	112
B-Exurban Dream	9	110
--ENTIRE SAMPLE----	9	100
M-Minority Bl Collr	8	93
E-Bright Beginnings	8	92
D-Rainbow Manors	8	89
K-Old Timers	7	85
G-New Collars	7	77
I-Minority Wh Collr	4	51
L-Downtown, P.G.	4	46

AWARD and/or TRANSFER	%	IDX
F-Rapid Development	40	132
B-Exurban Dream	39	128
I-Minority Wh Collr	39	127
G-New Collars	35	113
C-Aging White Collr	35	113
D-Rainbow Manors	33	107
J-Middle America	32	105
--ENTIRE SAMPLE----	31	100
M-Minority Bl Collr	30	98
K-Old Timers	26	84
A-Country Club	22	73
E-Bright Beginnings	20	65
H-Levittown, P.G.	18	58
L-Downtown, P.G.	12	40

STILL AT PGCC-SPR'88	%	IDX
E-Bright Beginnings	33	351
C-Aging White Collr	14	147
B-Exurban Dream	11	112
--ENTIRE SAMPLE----	9	100
F-Rapid Development	7	79
H-Levittown, P.G.	7	76
I-Minority Wh Collr	7	72
M-Minority Bl Collr	7	70
D-Rainbow Manors	7	70
L-Downtown, P.G.	4	43
K-Old Timers	3	29
A-Country Club	2	26
G-New Collars	0	0
J-Middle America	0	0

TRANSFD TO 4-YR INST.	%	IDX
B-Exurban Dream	38	141
F-Rapid Development	37	136
I-Minority Wh Collr	36	133
G-New Collars	34	127
C-Aging White Collr	31	116
D-Rainbow Manors	30	110
--ENTIRE SAMPLE----	27	100
J-Middle America	26	96
M-Minority Bl Collr	25	94
K-Old Timers	21	79
A-Country Club	19	70
E-Bright Beginnings	17	65
H-Levittown, P.G.	9	32
L-Downtown, P.G.	8	31

"PASS-THROUGHs"	%	IDX
B-Exurban Dream	42	168
C-Aging White Collr	39	156
I-Minority Wh Collr	36	144
D-Rainbow Manors	34	136
F-Rapid Development	33	132
G-New Collars	31	124
--ENTIRE SAMPLE----	25	100
M-Minority Bl Collr	22	88
K-Old Timers	19	76
J-Middle America	17	68
E-Bright Beginnings	12	49
A-Country Club	12	47
L-Downtown, P.G.	8	32
H-Levittown, P.G.	5	20

EXITED-NO GRAD./TRSF.	%	IDX
L-Downtown, P.G.	84	148
A-Country Club	75	133
H-Levittown, P.G.	75	133
K-Old Timers	71	126
J-Middle America	68	120
M-Minority Bl Collr	63	112
G-New Collars	63	111
--ENTIRE SAMPLE----	57	100
I-Minority Wh Collr	53	94
D-Rainbow Manors	52	92
F-Rapid Development	50	89
E-Bright Beginnings	47	83
B-Exurban Dream	38	67
C-Aging White Collr	36	64

Details of Re-Proportioning Methodology

The "Re-Proportioning" technique used in this study was an attempt to overcome what important sample distortion remained among our S.B.C.C. respondent group after major size reduction due to cluster-encoding because of the relative inefficiency of sample weighting in this circumstance. In essence, its works by the wholesale arithmetic forcing of sub-sample response proportions on Variable X into closer conformity with the known Variable X proportions for an unbiased comparison entire sample.

The exact steps were as follows:

1. S.B.C.C. survey outcome variables were re-worked into a single category variable which would allow the tracking of the entire branched sequence of possible outcomes:

Award Attainment/Transfer 4-Yr	No Award/Transfer 4-Yr
Award Attainment/Transfer 2-Yr	No Award/Transfer 2-Yr
Award Attainment/No Transfer	Still Enrolled/No Outcome
No Longer Enrolled/No Award/No Transfer	

With these seven categories, 100 percent of all students could be characterized as to latest academic career pattern at once.

2. Above category percentages were then calculated for:

   a. Weighted Original Whole Sample
   b. Re-Weighted Clusterized Whole Sample
   c. Each cluster sub-sample of (b).

3. (a) and (b) proportions were systematically compared by taking ratios A/B (re-proportioning coefficients):

Award-Trs4	4.80 % /	3.73 % =	1.287
Award-Trs2	.24 % /	.24 % =	.800
	.	.	.
	.	.	.
NoEnr-Aw-Tr	56.60 % /	65.97 % =	.858
	100.00 %	100.00 %	

4. These coefficients were then applied to corresponding categories in each cluster sub-sample -- coefficient x frequency. In this step frequency rather than percentage was used to facilitate the calculation of adjusted percentages which would re-sum to 100 percent -- e.g.:

$$\text{cat-1 \%} = 100*(c1*n1)/((c1*n1)+(c2*n2)+\ldots(c7*n7))$$
$$\text{cat-2 \%} = 100*(c2*n2)/((c1*n1)+(c2*n2)+\ldots(c7*n7))$$
$$\vdots$$
$$\frac{\text{cat-7 \%}}{100\ \%} = 100*(c7*n7)/((c1*n1)+(c2*n2)+\ldots(c7*n7))$$

for each cluster using the identical seven coefficients, where c=coefficient value and n=cluster category frequency.

5. Finally, create the six study indicators through category recombining:

Award Attainment	= Award-Trs4 + Award-Trs2 + Award-No Trs
Transfer 4-Year	= Award-Trs4 + No Aw-Trs4
Award/Transfer	= Award-Trs4 + Award-Trs2 + Award-No Trs + No Aw-Trs4
Pass-Through	= No Aw-Trs4 + No Aw-Trs2
Pipeline	= Still Enrld
Pass-Out	= No Enr-Aw-Tr

This technique, it must be said, does still involve some response distortion. What it does is trade off eliminated radical differences between biased sample proportions and known unbiased sample proportions for some added within sub-sample distortion. Remember that a constant coefficient is applied to all parallel proportions across sub-samples, which assumes -- mostly wrongly -- no across-sub-sample variation is proportional distortion. However, what usually can be assumed, as here, is no really important variation. Still, some new distortion is created, although less than is removed. The new distortion shows up as summed category percentages of less or more that 100 for each sub-sample, when coefficients are directly applied to category proportions rather than frequencies as in Step 4. In our case summed percentages across clusters ranged between 94-107, indicating relatively small additional distortion. Step 4's arithmetic has the same effect as would directly adjusting percentage sums to force a 100 percent total -- for example:

$$5+10+15+20+25+30 = 105$$

$$5/105 = 4.76$$
$$10/105 = 9.52$$
$$\vdots$$
$$30/105 = 28.57$$
$$\overline{100.00}$$

ERIC Clearinghouse for Junior Colleges

MAY 3 1991